Proceedings IEEE Virtual Reality '99, pages 96–103, IEEE Computer Society Press, 1999.
*Awarded **Best Paper** at VR'99 Conference.*

User-Centered Design and Evaluation of a Real-Time Battlefield Visualization Virtual Environment

Deborah Hix[1], J. Edward Swan II[2], Joseph L. Gabbard[3], Mike McGee[1], Jim Durbin[2], Tony King[2]

[1]Virginia Polytechnic Institute and State University, Blacksburg, VA
[2]The Naval Research Laboratory, Washington, D.C.
[3]Virtual Prototyping and Simulation Technologies, Inc., Blacksburg, VA

Abstract

The ever-increasing power of computers and hardware rendering systems has, to date, primarily motivated the creation of visually rich and perceptually realistic virtual environment (VE) applications. Comparatively very little effort has been expended on the user interaction components of VEs. As a result, VE user interfaces are often poorly designed and are rarely evaluated with users. Although usability engineering is a newly emerging facet of VE development, user-centered design and usability evaluation in VEs as a practice still lags far behind what is needed.

This paper presents a structured, iterative approach for the user-centered design and evaluation of VE user interaction. This approach consists of the iterative use of expert heuristic evaluation, followed by formative usability evaluation, followed by summative evaluation. We describe our application of this approach to a real-world VE for battlefield visualization, describe the resulting series of design iterations, and present evidence that this approach provides a cost-effective strategy for assessing and iteratively improving user interaction design in VEs. This paper is among the first to report applying an iterative, structured, user-centered design and evaluation approach to VE user interaction design.

Keywords: user-centered design, user interfaces, user interaction, user assessment, usability engineering, usability evaluation, virtual environments, virtual reality, expert heuristic evaluation, formative evaluation.

1 Introduction and Related Work

Despite the ever-increasing power of computers and hardware rendering systems, the user interaction components of VE applications are often poorly designed and are rarely evaluated with users. The vast majority of VE research and design effort has been on the development of visual quality and rendering efficiency. As a result, many visually compelling VEs are difficult to use and are, therefore, non-productive for their users. While these VEs might make good entertainment applications, their usability problems prevent them from being useful for efficiently solving real-world problems.

Usability engineering [10] and user-centered design [11] are newly emerging facets of VE design and evaluation. VE designers and developers are becoming aware of traditional human-computer interface (HCI) usability research and are beginning to apply and expand upon those methods for VEs. A few efforts have been reported to date; however, user-centered design and usability evaluation in VEs as a practice still lags far behind what is needed.

One reported work on user-based evaluation in VEs is Bowman et al. [1], who investigated an aspect of navigation in VEs and present a framework for evaluating travel (viewpoint motion control). The framework supports a methodology for evaluating different VE travel techniques and for appropriately matching travel techniques with virtual applications. Several aspects, or quality factors, were identified as being important to travel: speed, accuracy, spatial awareness, ease of learning, information gathering, presence, and user comfort. The authors acknowledge that task-related factors (task, environment, user, and system characteristics) can have a greater impact on quality factor performance than the travel technique selected. The evaluation methodology described is intended to be generalizable to a variety of VEs.

Salzman et al. [14] discuss how usability engineering methods shaped iterative development of a VE designed for educating students on various concepts associated with Newton's laws of physics. The goal of the design process was to develop a usable and educational virtual world. The authors applied usability evaluation to identify and refine early system weaknesses across three premises: usability, learning, and learning vs. usability. Both potential users (high school students) and experts in the field (physics professors) participated in the formative evaluations, which resulted in changes that improved the final VE user interaction.

Other research that has reported a limited element of usability evaluation includes a study of haptic interfaces [6], and an investigation of spatial input devices [7]. In addition, Stuart [16] describes basic methods for evaluating general usability components of VEs.

While these efforts provide insights about usability issues of specific VE technology, most do not provide sufficient breadth for large, complex VE design and assessment. Gabbard and Hix [4] propose a framework of usability characteristics structured to support usability engineering of VEs. They present a methodology for approaching design and assessment of VE user interfaces, which employs a top-down, step-wise refinement of VE usability space. This framework was used during evaluation of the battlefield visualization VE described herein (see Section 4.3 and Section 5).

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE **1999**	2. REPORT TYPE	3. DATES COVERED **00-00-1999 to 00-00-1999**

4. TITLE AND SUBTITLE **User-Centered Design and Evaluation of a Real-Time Battlefield Visualization Virtual Environment**	5a. CONTRACT NUMBER 5b. GRANT NUMBER 5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER 5e. TASK NUMBER 5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Naval Research Laboratory,4555 Overlook Ave. SW,Washington,DC,20375**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR’S ACRONYM(S) 11. SPONSOR/MONITOR’S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT
The ever-increasing power of computers and hardware rendering systems has, to date, primarily motivated the creation of visually rich and perceptually realistic virtual environment (VE) applications. Comparatively very little effort has been expended on the user interaction components of VEs. As a result, VE user interfaces are often poorly designed and are rarely evaluated with users. Although usability engineering is a newly emerging facet of VE development, user-centered design and usability evaluation in VEs as a practice still lags far behind what is needed. This paper presents a structured, iterative approach for the user-centered design and evaluation of VE user interaction. This approach consists of the iterative use of expert heuristic evaluation, followed by formative usability evaluation, followed by summative evaluation. We describe our application of this approach to a real-world VE for battlefield visualization, describe the resulting series of design iterations, and present evidence that this approach provides a cost-effective strategy for assessing and iteratively improving user interaction design in VEs. This paper is among the first to report applying an iterative structured, user-centered design and evaluation approach to VE user interaction design.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**Same as Report (SAR)**	**8**	

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std Z39-18

Proceedings IEEE Virtual Reality ’99, pages 96–103, IEEE Computer Society Press, 1999.
*Awarded **Best Paper** at VR’99 Conference.*

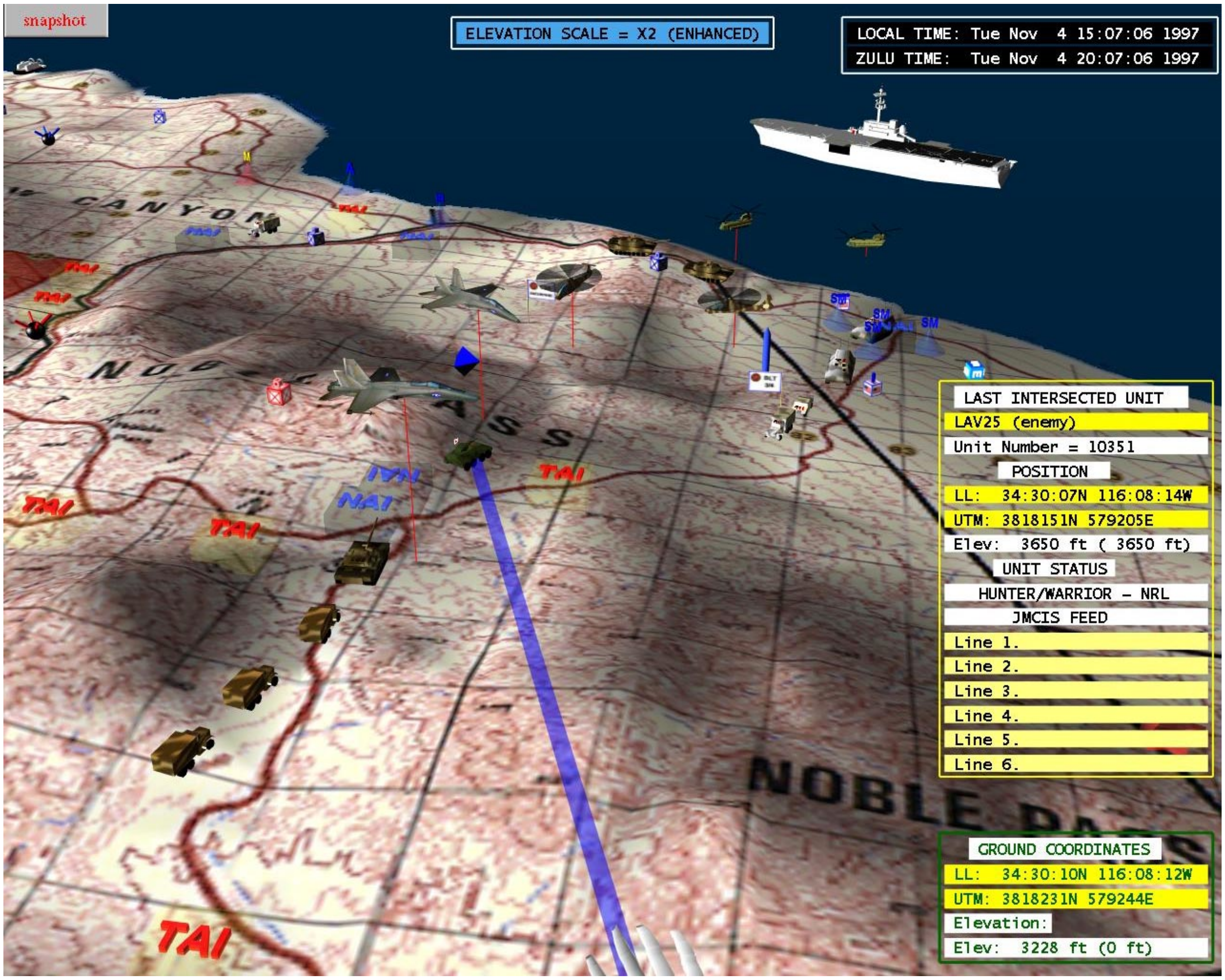

Figure 1: Screen shot from the Dragon battlefield visualization virtual environment.

Personnel at the Naval Research Laboratory’s (NRL) Virtual Reality Lab have developed a VE for battlefield visualization, called Dragon (Figure 1) [3], which is implemented on a Responsive Workbench [9, 13]. The responsive workbench provides a natural metaphor for visualizing and interacting with three-dimensional computer-generated scenes using a familiar tabletop environment. Applications in which several users collaborate around a workspace, such as a table, are excellent candidates for the workbench. Researchers from NRL, collaboratively with researchers from Virginia Tech, are empirically studying the most important usability parameters of an effective VE user interface for Dragon.

In the next section, we discuss battlefield visualization in general, and we describe the Dragon battlefield visualization VE. In Section 3, we discuss three important usability evaluation methods that can be profitably applied to VEs: expert heuristic evaluation, formative evaluation, and summative evaluation. In Section 4 we present our methodological approach for applying expert heuristic and formative evaluation methods to Dragon’s design and evaluation, and in Section 5 we describe and discuss the design iterations that resulted from using this approach. In Section 6, we discuss lessons learned from this work, including evidence that our structured approach provides a cost-effective strategy for assessing and iteratively improving user interaction designs in VEs. We conclude with ideas for future work, particularly summative evaluation.

2 The Dragon Real-Time Battlefield Visualization Virtual Environment

2.1. Battlefield Visualization and Dragon

For decades, battlefield visualization has been accomplished by placing paper maps of the battlespace under sheets of acetate. As intelligence reports arrive from the field, technicians use grease pencils to mark new information on the acetate. Commanders then draw on the acetate to plan and direct various battlefield situations. Thus, the map and acetate together present a visualization of the battlespace. Using maps and overlays can take several hours to print, distribute, and update. Historically (before high-quality paper maps) these same operations were performed on a *sandtable* (a box filled with sand shaped to replicate the battlespace terrain). Commanders moved around small physical replicas of battlefield objects to direct battlefield situations. Currently, the fast-changing modern battlefield pro-

duces so much time-critical information that these cumbersome, time-consuming methods are inadequate for effectively visualizing the battlespace.

In Dragon, the workbench provides a three-dimensional display for observing and managing battlespace information shared among commanders and other battle planners. Visualized information includes a high-resolution terrain map; entities representing friendly, enemy, unknown, and neutral units; and symbology representing other features such as obstructions or key battle objectives. Dragon receives electronic intelligence feeds that provide constantly updated, displayable information about each entity's status, including position, speed, heading, damage condition, and so forth. Users can navigate to observe the map and entities from any angle and orientation, and can query and manipulate entities.

2.2. Design of User Interaction in Dragon

Early in Dragon's development, we developed and assessed three general interaction methods for the workbench, any of which could have been used to interact with Dragon: hand gestures using a pinchglove [12], speech recognition, and a hand-held flightstick. Although an interesting possibility for VE interaction, we found speech recognition still too immature for battlefield visualization, and we found the pinchglove to be fragile, time-consuming to pass from user to user, and limiting in that it requires right-handed users whose hands are approximately the same size. In contrast, we found the hand-held flightstick to be robust, easily handed from user to user, and applicable to both right- and left-handed users.

Based on these observations, we modified a three-button game flightstick by removing its base and placing a six degree-of-freedom position sensor inside. We tracked the flightstick's position and orientation relative to an emitter located on the front center of the workbench. We accomplished VE interaction with a *virtual laser pointer* metaphor: a laser beam appears to come out of the flightstick, allowing interaction with the terrain or object that the beam intersects.

Early in its development, when very little usability evaluation had been performed, Dragon was demonstrated as a prototype system at two different military exercises. In both demonstrations, an objective was a proof-of-concept for using a workbench-based battlefield visualization tool. Feedback from both civilian and military VIPs indicated that users found Dragon's battlespace visualization to be more effective and efficient than the traditional method of maps, acetate, and grease pencils. Following these successful demonstrations, we began intensive usability evaluations and iterations of Dragon's user interface.

3 Usability Evaluation Methods

User-based evaluation is an essential component of developing any interactive application, and is especially important for applications as complex and innovative as VEs. Three kinds of usability evaluation are particularly appropriate: expert heuristic evaluation, formative evaluation, and summative evaluation. We performed the first two types extensively during Dragon's development (Sections 4 and 5), and have plans for the third type (Section 6).

Expert heuristic evaluation [10] is a type of analytical evaluation in which an expert in user interaction design assesses a particular user interface by determining what usability design guidelines it violates and supports. Then, based on these findings, especially the violations, the expert makes recommendations for changes to improve the design. In the case of VEs, this is particularly challenging because there are so few guidelines that are specific to VE user interfaces. Thus, users are not directly involved in expert heuristic evaluation. Typically, this type of usability evaluation is more effective if the experts are not also developers of the user interaction design being evaluated. This was our situation: the first three authors of this paper, who were not involved with development of Dragon, did much of the expert heuristic evaluation described in Section 4.3.

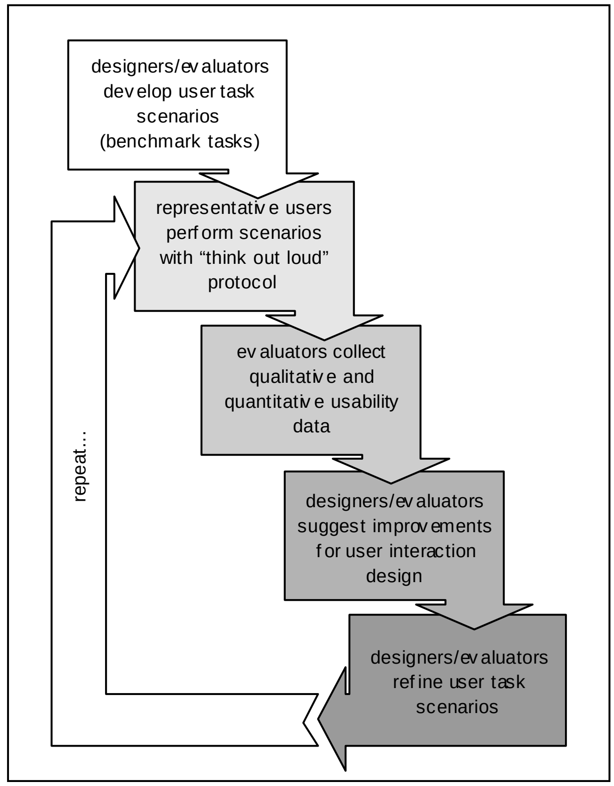

Figure 2: Formative evaluation process.

Formative evaluation [8] is a type of empirical, observational assessment *with users* that begins in the earliest phases of user interaction design and continues throughout the entire life cycle. Formative evaluation produces both qualitative (narrative) and quantitative (numeric) results. The purpose of formative evaluation is to iteratively and quantifiably assess and improve the user interaction design.

An important point to note in the formative evaluation process, shown in Figure 2, is that both qualitative and quantitative data are collected from representative users during their performance of task scenarios. Developers often have the false impression that usability evaluation is something rather warm and fuzzy, with no "real" process and collecting no "real" data. Quite the contrary is true; experienced usability evaluators collect large volumes of both qualitative data and quantitative data.

Qualitative data are typically in the form of *critical incidents* [5, 8]. A critical incident occurs while a user is performing task scenarios, and is an event that has a significant effect, either positive or negative, on user task performance or user satisfac-

tion with the interface. Events that affect user performance or satisfaction therefore have an impact on usability. Typically, a critical incident is a problem that a user encounters (e.g., an error, being unable to complete a task scenario, confusion, etc.). Section 5 describes the major design iterations that resulted from hundreds of critical incidents, which we collected during our formative evaluation studies.

Quantitative data are generally related, for example, to how long it takes and the number of errors committed while a user is performing task scenarios. These data are then compared to appropriate baseline metrics. Quantitative data generally indicate that a problem has occurred; qualitative data indicate where (and sometimes why) it occurred.

Collection of both these types of data is an important part of the formative evaluation process. While we focused largely on qualitative, critical incident data, we also collected some quantitative data.

Summative evaluation [8], in contrast, is an empirical assessment with users of an interaction design in comparison with other interaction designs for performing the same user tasks. Summative evaluation is typically performed when there are some more-or-less "final" versions of the interaction designs, and it yields primarily quantitative results. The purpose of summative evaluation is to statistically compare user performance with different interaction designs, for example, to determine which one is better, where "better" is defined in advance. Summative evaluations of Dragon are planned (Section 6).

Best guesses about an interaction design are substantiated or refuted by many tight, short cycles of heuristic and formative evaluation. During the design and assessment of the Dragon VE user interface, we performed numerous cycles of heuristic and formative evaluation—some as short as a few minutes (these were the really bad designs!), others lasting several hours. Evolution of essentially all decisions about design details came from many rounds of evaluation. As discussed in the following sections, from the heuristic and formative evaluations we have greatly improved Dragon's user interaction design, and are now planning a summative study.

4 Method: Application of Design and Evaluation Methods

4.1 Focus on Navigation

During our early demonstrations and evaluations, we observed that *navigation* — how users manipulate their viewpoint to move from place to place in a virtual world (in this case, the map for battlefield visualization) — profoundly affects all other user tasks. If a user cannot successfully navigate in a virtual world, then other user tasks (e.g., involving specific objects or groups of objects) simply cannot be performed. A user cannot query an object if the user cannot navigate through the virtual world to get to that object. Although we performed a user task analysis before our heuristic and formative studies, these studies corroborated our task analysis and our expectations of what tasks are most important.

Further, our observational studies revealed several other generic tasks performed by users of battlefield visualization VEs, including object manipulation, object selection, object querying, query response, and object aggregation. These user tasks will become the focus of possible future research for us and for others. Again, without having performed the expert and formative usability evaluations, we would only be able to guess at our assumptions about user tasks.

4.2 Methodology

We used the basic Dragon application as an instrumentable testbed, modified as needed for our heuristic and formative usability evaluation purposes. We performed extensive evaluations over a nine-month period, using anywhere from one to three users for each cycle of evaluation. From a single evaluation session, we often uncovered design problems so serious that it was pointless to have a different user attempt to perform the scenarios with the same design. So we would iterate the design, based on our observations, and begin a new cycle of evaluation. We went through four major cycles of iteration (Section 5).

Based on our task analysis and early evaluations, we created a set of scenarios comprised of benchmark user tasks, carefully considered for coverage of specific issues related to navigation. For example, some of the tasks exploited an ego-centric (user moves through world) navigation metaphor while others exploited an exo-centric (user moves the world) navigation metaphor (see Section 5). Some scenarios exercised various navigation tasks (i.e., degrees of freedom: pan, zoom, rotate, heading, pitch, roll) throughout the virtual map world. Other scenarios served as primed exploration or non-primed searches [2], while still others were designed to evaluate rate control versus position control in the virtual world. We thoroughly pre-tested and "debugged" all scenarios before presenting them to users during an evaluation session.

4.3 Expert Heuristic Evaluations

During our expert heuristic evaluations, various user interaction design experts worked alone or collectively to assess the evolving user interaction design for Dragon. In our earliest heuristic evaluations, the experts did not follow specific user task scenarios per se, but engaged simply in "free play" with the user interface. All experts knew enough about the purpose of Dragon as a battlefield visualization VE to explore the kinds of tasks that would be most important for users of Dragon. During each heuristic evaluation session, one person was typically "the driver," holding the flightstick and generally deciding what and how to explore in the application. One and sometimes two other experts were observing and commenting. Much discussion occurred during each session.

As mentioned earlier, the first three authors of this paper were often the experts assessing the current design. Their assessment and discussions were guided largely by their own knowledge of interaction design for VEs, and, more formally, by a framework for usability characteristics of VEs [4], discussed in Section 1. This framework provided a more structured means of evaluation than merely wandering around at random in the application, and provided guidance on how to make modifications to improve discovered design guideline violations. The major design problems uncovered by the expert heuristic evaluations were: 1) poor mapping of navigation tasks (e.g., pan, zoom, pitch, heading) to flightstick buttons, 2) missing functionality (e.g., exo-centric rotate, terrain following), 3) problems with damping of map movement in response to flightstick movement, and 4) graphical and textual feedback to the user about the current navigation task (e.g., pan, zoom, etc.). These problems, and

how we addressed them, are discussed further in Section 5. After our cycles of expert heuristic evaluation had revealed and remedied as many design flaws as possible, we moved on to formative evaluations.

4.4 Formative Evaluations

During each of six formative evaluation sessions, we followed a formal protocol of welcoming the user, giving them an overview of the evaluation about to be performed, and then explaining the responsive workbench and the Dragon application. We were careful to *not* explain too many details of the Dragon interaction design, since that was what we were evaluating. Then the user was asked to play with the flightstick to figure out which button activated which navigation task (e.g., pan, zoom, etc.). We timed each user as they attempted to determine this, and took notes on comments they made and any critical incidents that occurred. Once a user had successfully figured out how to use the flightstick, we began having them perform the scenarios. If about 15 minutes passed without a user figuring out the flightstick and its buttons (this happened in only one case), we filled in details that they had not yet determined and moved on to scenarios.

Time to perform the set of scenarios ranged from about 20 minutes to more than an hour. We timed user performance of individual tasks and scenarios, and counted errors they made during task performance (quantitative data). A typical error was moving the flightstick in the wrong direction for the particular navigation metaphor (exo-centric or ego-centric) that was currently in use. Other errors involved simply not being able to maneuver the map (e.g., to rotate it) and persistent problems with mapping navigation tasks to flightstick buttons. Again, these are discussed further in Section 5. We also carefully noted critical incidents, especially related to errors, and constructive comments users made about the design (qualitative data).

During each session, we had at least two and often three evaluators present: one was the "leader" who ran the session and interacted with the user; the other one or two evaluators recorded timings, counted errors, and collected qualitative data. While both the expert heuristic evaluation sessions and the formative evaluation sessions were personnel-intensive (with two or three evaluators involved), we found that the quality and amount of data collected by multiple evaluators greatly outweighed the cost of those evaluators. After each session, we analyzed both the quantitative and qualitative data, and based our next iteration on our results, as explained in the next section.

5 Results: Iterations of the Dragon User Interaction Design

Table 1 summarizes the four major iterations of the Dragon user interaction design over an approximately one-year period. It gives a high-level description of each iteration (including both visual and flightstick characteristics), and indicates the major usability findings for each iteration. (Space does not permit us to explain all the information in this table in detail.) Our findings, shown in rows of the table, fell into four categories:

General Description. For each iteration, we give a brief descriptive title in the top four cells of Table 1. A general description of each iteration's most salient features is shown beneath, along with the approximate date when the iteration was completed.

Interaction Description. This category describes some specifics of how a user interacts with each design iteration. We experimented extensively with variants of two different navigation metaphors (described below): exo-centric and ego-centric. We visualized the virtual laser pointer (see Section 2.2) by drawing a beam coming out of the flightstick and intersecting the environment. In the first ("Virtual Sandtable") iteration, we also drew a skeletal hand "holding" the beam to visualize the user's hand (lower edge of Figure 1). This category of Table 1 also shows the degrees of freedom used by the flightstick tracker.

Device Description. This category defines the mappings from the three flightstick buttons (left, right, and trigger) to degrees of freedom; examples are explained below.

Evaluation Results. This category indicates which evaluations were performed on each iteration, and summarizes major strengths and major flaws of each. The last row of Table 1 summarizes our user interaction design modification recommendations to Dragon's programmers.

During early design, we implemented two navigation metaphors: exo-centric (or map-centric) and ego-centric (or user-centric). An *exo-centric navigation metaphor* is based on how a user would interact with a real physical map on a table. Different buttons are used for navigation tasks such as pan, zoom, and pitch. The map mimics the motion of the flightstick, so that the map acts as if it is stuck to the laser beam; user movement of the flightstick in any direction causes the map to move in that same direction. The magnitude of a user's gesture controls the distance of the map's movement in the virtual world (this is also called *zero-order motion*). This means that, for example, when panning from side to side of a zoomed-in map, a user must make repeated panning gestures, each of which translates the map a distance equivalent to the length of the user's gesture.

An *ego-centric navigation metaphor* is loosely based on the concept of a user flying above the map as if in an airplane. Various button combinations are again used for navigation tasks. The magnitude of a user's gesture controls the velocity of the map's movement (also called *first-order motion*); for example, a user can fly from one side to the other of a zoomed-in map with a single gesture.

The first iteration, "Virtual Sandtable", was based on the sandtable concept briefly described in Section 2.1, and was the version demonstrated in the military exercises mentioned in Section 2.2. So in addition to expert heuristic evaluation, we had feedback from the demonstrations. A key finding of this iteration was that users wanted a terrain-following capability, allowing them to "fly" over the map. Based on observations of users interacting with maps in a combat center, we had initially thought that a battlespace visualization application only required an exo-centric navigation metaphor. In reality, the workbench-based Dragon creates a very rich environment, in which users can do much more than just move a map. They can actually experience the environment by visually sizing up terrain features, entity placement, fields of fire, lines of sight, and so forth. Exo-centric navigation worked well when globally manipulating the environment and conducting operations on large-scale units. However, for small-scale operations, users wanted the "fly" capability. The logical approach to designing this into Dragon was an ego-centric flying capability. We found that the mapping of flightstick buttons to navigation tasks shown in Table 1 (i.e., trigger and left button pressed simultaneously produced

combined pan and zoom; trigger and right button together produced combined heading and pitch) worked well for users.

In designing the second iteration, "Point and Go," we used the framework of usability characteristics of VEs [4] (see Section 1) to suggest various possibilities for an ego-centric navigation metaphor design, such as WIM [15] and eye-in-hand [17]. We ultimately designed a "point and go" metaphor, in which we attempted to avoid having different modes (and buttons) for different navigation tasks (pan, zoom, etc.) because of known usability problems with moded interaction. Further, we based this decision on how a person often navigates to an object or location in the real world; namely, they point (or look) and then go (move) there. Our reasoning was that adopting this same idea to ego-centric navigation would simplify the design and at least loosely mimic the real world. So in this iteration, a user simply pointed the flightstick toward a location or object of interest, and pressed the trigger to fly there. We found through our expert heuristic evaluation that the single gesture to move about was not powerful enough to support the diverse, complicated navigation tasks inherent in Dragon. Furthermore, a single gesture meant that all degrees of freedom were controlled by that single gesture. This resulted in, for example, unintentional rolling when a user only wanted to pan or zoom. Essentially, we observed a *control* versus *convenience* trade-off. Many navigation tasks (modes) were active simultaneously, which was convenient but difficult to physically control. With separate tasks (modes), there was less convenience but physical control was easier because degrees of freedom were more limited in each mode. In addition to these serious problems, we found that users wanted to rotate around an object, such as to move completely around a tank and observe it from all sides. This indicated that Dragon needed an exo-centric rotate ability, which was added. This interesting finding showed that neither a pure ego-centric nor a pure exo-centric metaphor was desirable; each metaphor has aspects that are more or less useful depending on user goals.

In the third iteration, "Modal," we went from the extreme of all navigation tasks coupled on a single button to a rather opposite design in which each navigation task was a separate mode. Specifically, as a user clicked the left or right flightstick button, Dragon cycled successively through the tasks of pan, zoom, pitch, heading, and exo-centric rotate. Once a user had cycled to the desired task, it was enabled and thus accessible from the trigger, and the task name appeared in a small textual indicator. We observed that, as we expected, it was very cumbersome for users to always have to cycle between modes, and it was obvious that we still had not achieved a compromise between convenience and control. Again using the framework of usability characteristics of VEs [4] for guidance, for our fourth iteration of the Dragon interaction design, "Integrated Navigation," we decided to couple pan and zoom onto the flightstick trigger, pitch and heading onto a single button, and exo-centric rotate and zoom onto the third flightstick button, as indicated in Table 1. Our fourth generation design appears to have achieved the desired convenience versus control compromise. In our final evaluation studies, we found that at last we had a design for navigation that seemed to work well for most users. The only problem we observed was minor: damping of map movement was too great and needed some adjustment, which we made.

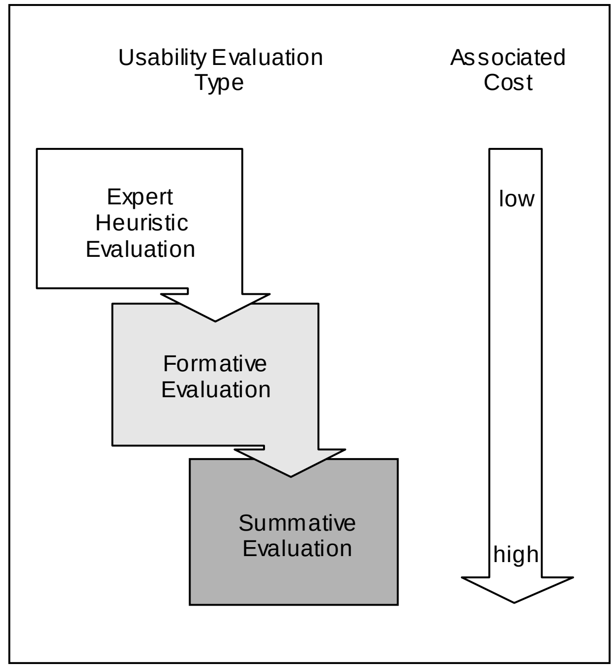

Figure 3: Types of usability evaluation and their cost.

6 Lessons Learned and Future Work

A key finding of our research is the successful progression from heuristic to formative to summative evaluations as a very cost-effective strategy for assessing and improving a user interaction design. Far too often, summative studies are conducted on applications whose interaction design has had little or no heuristic or formative evaluation. This situation is unfortunate because it is often the case that very expensive summative evaluations are comparing "good apples" with "bad oranges". That is, the differences between two interaction designs may occur because one design is inherently better, in terms of usability, than the other. If both designs have been heuristically and/or formatively evaluated, then experimenters can have confidence that the interaction designs are essentially equivalent in terms of their usability: they will be comparing "good apples" to "good oranges". And it is therefore much more likely that any differences found in a summative comparison are truly due to differences in the nature of the applications, and not in their user interaction designs per se.

Further, the cost of performing these three types of evaluations typically ranges from lowest for expert heuristic evaluations to highest for summative evaluations, as shown in Figure 3. So if expert heuristic evaluations are not performed prior to formative evaluations, the formative evaluations will typically take longer and require more users, and yet reveal many of the same usability problems that could generally have been discovered by less expensive heuristic evaluations. Thus, expert heuristic evaluations can reduce the cost of formative studies, and formative studies produce interaction designs that are truly comparable in summative studies for uncovering differences between applications.

Proceedings IEEE Virtual Reality '99, pages 96–103, IEEE Computer Society Press, 1999.
*Awarded **Best Paper** at VR'99 Conference.*

	Virtual Sandtable	Point & Go	Modal	Integrated Navigation
General Description	sandtable metaphor	one gesture moves anywhere on map	all navigation tasks separated into discrete modes	modes mapped to all three flightstick buttons
Approximate Date	June 1997	November 1997	January 1998	April 1998
Interaction Description				
Navigation Metaphor	exo-centric (map-centric)	ego-centric (flying)	primarily ego-centric, except for exo rotate	primarily ego-centric, except for exo rotate
Laser Pointer Visual Representation	laser pointer & skeleton hand	laser pointer	laser pointer	laser pointer
Supported Degrees of Freedom	x, y, z, heading, pitch	x, y, z, heading, pitch, roll	x, y, z, heading, pitch	x, y, z, heading, pitch
Device Description				
Button Mappings	trigger & left→pan & zoom trigger & right→heading & pitch	trigger→pan & zoom & pitch & heading & roll	left and right buttons cycle modes: pan, zoom, pitch, heading, exo rotate	trigger→pan & zoom left→pitch & heading right→exo rotate & zoom
Evaluation Results				
Evaluations Performed	heuristic	heuristic	heuristic and formative	heuristic and formative
Major Strengths of Iteration	• easy to pan/zoom • good for overview tasks	• modeless navigation	• easy navigation to any location with single mode	• easy navigation to any location • easy to switch between navigation tasks
Major Flaws of Iteration	• skeleton hand orientation did not match user hand orientation • terrain following difficult • pan gesture parallel to floor not workbench screen	• hard to travel to non-visible location on map • could travel underneath map • trigger overloaded with too many degrees of freedom • many navigation tasks resulted in unintentional rolling	• too cumbersome to switch between modes	• too much damping; user movement too slow • zoom gesture parallel to workbench screen, not floor
Recommendations to Programmers for Interaction Design Changes	• support terrain following	• fine-tune damping and acceleration • add collision detection with map • remove ability to roll • add exo-centric rotation	• couple modes so that only three navigation modes because then can map to three buttons on flightstick • couple pitch and heading • couple pan and zoom	• fine-tune damping and acceleration

Table 1: Major iterations of Dragon user interaction design.

Our future work will focus on summatively evaluating our current navigation design. During our expert heuristic and formative evaluations, we discovered many different variables that affect navigation usability in VEs. We have narrowed this (initially large) list to five variables, based on the framework of usability characteristics [4], our observations during heuristic and formative evaluations, and our expertise in VE interaction design. We feel these five variables have the greatest effect on navigation, and are therefore the most important candidates for summative evaluations:

1) *navigation metaphor* (ego- vs. exo-centric),
2) *gesture control* (controls rate vs. controls position),
3) *visual presentation device* (workbench, desktop, CAVE™),
4) *head tracking* (present vs. not present), and
5) *stereopsis* (present vs. not present).

An expected result of these planned studies is empirically determined guidelines for navigation design in VEs.

To summarize, our research has produced results at three levels:

1) important navigation improvements in Dragon,
2) recommendations for navigation design in VEs, especially workbench-based VEs, and
3) evidential substantiation of a structured approach for user-centered design and evaluation of VEs.

This paper is one of the first to report using expert heuristic evaluation followed by formative usability evaluation as a structured approach to the iterative, user-centered design and evaluation of VE user interaction components. Our use of this approach with a real-world battlefield visualization VE has resulted in a VE for which we have empirical evidence of effectiveness and usability.

Acknowledgements

Many people have contributed to Dragon's development and therefore to this reported work. Eddy Kuo, in many late nights, went far beyond the normal call of duty to get Dragon ready for evaluation. Brad Colbert and Chris Scannell made many of the recommended changes throughout development. Other developers included John Crowe, Josh Davies, Bob Doyle, Rob King, Greg Newton, and Josh Summers.

At NRL, Dr. Larry Rosenblum and Dave Tate gave leadership, inspiration, and guidance to this project. Dr. Jim Templeman and Linda Sibert of NRL and Dr. Bob Williges of Virginia Tech provided valuable suggestions.

This research was funded by the Office of Naval Research under Program Managers Helen Gigley and Paul Quinn, and by the United States Marine Corps. We would like to thank Dr. Gigley for her continued support of an on-going synergistic collaboration in human-computer interaction research between Virginia Tech and NRL over the past several years.

References

[1] Bowman, D, Koller, D, and Hodges, LF. (1998). "A Methodology for the Evaluation of Travel Techniques for Immersive Virtual Environments". *Virtual Reality: Journal of the Virtual Reality Society*, 3, pp. 120–131.

[2] Darken, RP and Sibert, JL. (1996). "Wayfinding Strategies and Behaviors in Large Virtual Environments", In Proc. *Human Factors in Computing Systems (CHI '96)*, pp. 142–149.

[3] Durbin, J, Swan II, JE, Colbert, B, Crowe, J, King, R, King, T, Scannell, C, Wartell, Z, and Welsh, T. (1998). "Battlefield Visualization on the Responsive Workbench", In Proc. *IEEE Visualization '98*, IEEE Computer Society Press, pp. 463–466.

[4] Gabbard, JL. *A Taxonomy of Usability Characteristics in Virtual Environments*. (1997). M.S. Thesis, Department of Computer Science and Applications, Virginia Polytechnic Institute and State University. Available from

http://www.theses.org/vt.htm and

http://www.vpst.org/jgabbard/ve/framework/.

Also Gabbard JL and Hix, D. (1998). "Usability Engineering for Virtual Environments through a Framework of Usability Characteristics", submitted to *Presence: A Journal for Teleoperators and Virtual Environments*.

[5] del Galdo, EM, Williges, RC, Williges, BH, and Wixon, DR. (1986). "An Evaluation of Critical Incidents for Software Documentation Design", In Proc. *30th Annual Human Factors and Ergonomics Society Conference*, pp. 19–23.

[6] Hannaford, B and Venema, S. (1995). "Kinesthetic Displays for Remote and Virtual Environments", In W Barfield and T Furness (eds.) *Virtual Environments and Advanced Interface Design*. Oxford University Press.

[7] Hinckley, K, Pausch, R, Goble, JC, and Kassell, NF. (1994). "Design Hints for Spatial Input", In Proc. *ACM Symposium on User Interface Software & Technology (UIST '94)*, pp. 213–222.

[8] Hix, D and Hartson, HR. (1993). *Developing User Interfaces: Ensuring Usability Through Product & Process*. John Wiley & Sons, Inc.

[9] Kruger, W, Bohn, CA, Frohlich, B, Schuth, H, Strauss, W, and Wesche, G. (1995). "The Responsive Workbench: A Virtual Work Environment", *IEEE Computer*, 28(7), pp. 42–48.

[10] Nielson, J. (1993), *Usability Engineering*, Academic Press.

[11] Norman, DA and SW Draper, Eds. (1986). *User Centered System Design*. Lawrence Erlbaum Associates.

[12] Obeysekare, U, Williams, C, Durbin, J, Rosenblum, L, Rosenberg, R, Grinstein, F, Ramamurthi, R, Landsberg, A, and Sandberg, W. (1996). "Virtual Workbench—A Non-Immersive Virtual Environment for Visualizing and Interacting with 3D Objects for Scientific Visualization", In Proc. *IEEE Visualization '96*, IEEE Computer Society Press, pp. 345–349.

[13] Rosenblum, L, Durbin, J, Doyle, R, and Tate, D. (1998). "The Virtual Reality Responsive Workbench", *Virtual Worlds on the WWW, Internet, and Networks*, IEEE Computer Society Press.

[14] Salzman, MC, Dede, C, and Loftin, RB. (1995). "Usability and Learning in Education Virtual Realities", In Proc. *39th Annual Human Factors and Ergonomics Society Conference*, pp. 486–490.

[15] Stoakley, R, Conway, MJ, and Pausch, R. (1995). "Virtual Reality on a WIM: Interactive Worlds in Miniature", In Proc. *Human Factors in Computing Systems, (CHI '95)*, pp. 265–272.

[16] Stuart, R. (1996). *The Design of Virtual Environments*. McGraw-Hill.

[17] Ware, C and Osborne, S. (1990). "Exploration and Virtual Camera Control in Virtual Three Dimensional Environments", *Computer Graphics*, 24(2), pp. 175–183.